Operational Recommendations for Capturing History and Infusing Data Science (ORCHIDS)

Preparing your mission data for future analysis

Jack Lightholder

Lukas Mandrake

James Montgomery

Sarah Myers

Kevin Shannon

Brian Kahovec

Masha Liukis

JPL
Jet Propulsion Laboratory
California Institute of Technology

Motivation: Trapped Mission Data, Limited Analysis

Operating
Spacecraft

Science + Engineering

Ground
Station

- Spacecraft downlink both Science and Engineering Data

- Traditional Operations treat these flows very differently

JPL

Motivation: Trapped Mission Data, Limited Analysis

Operating Spacecraft

Ground Station

Science Data

- Science data is destined for exploration
- L0-L3 processing prepares data for analysis
- Standardized archives (DAAC's) serve data
- Search abilities permit researcher access
- Quality assessment and metadata captured
- Excellent document is assumed and required

Wild & Wooly
L0 Raw Data

Interpretable Calibrated
L1 Unitful Observations

Usable Traceable Quality Assessed
L2 Science Estimations

Harmonized Common Grid Common Time Filtered
L3 Harmonized Products

Easy Search Accessible Documented Common Formats
Standardized Archive

Science Pipeline is Healthy & Optimized for Analysis
(because that's the point of science data)
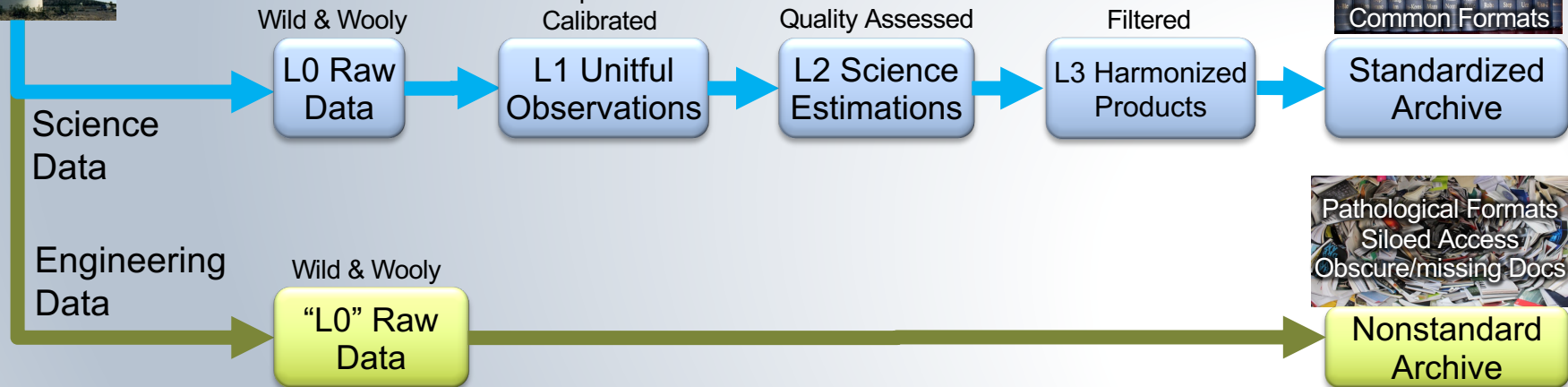
JPL

Motivation: Trapped Mission Data, Limited Analysis

Operating Spacecraft

- Engineering data is for "expert consumption only"
- Shallow or emergency analysis assumed
- Wake the person who understands this stuff!
- New personnel overwhelmed with missing details
- Even basic analysis is stymied,
  let alone data-driven science

Ground Station

Science Data

Engineering Data

Wild & Wooly

L0 Raw Data

Interpretable Calibrated

L1 Unitful Observations

Usable Traceable Quality Assessed

L2 Science Estimations

Harmonized Common Grid Common Time Filtered

L3 Harmonized Products

Easy Search Accessible Documented Common Formats

Standardized Archive

Wild & Wooly

"L0" Raw Data

Pathological Formats Siloed Access Obscure/missing Docs
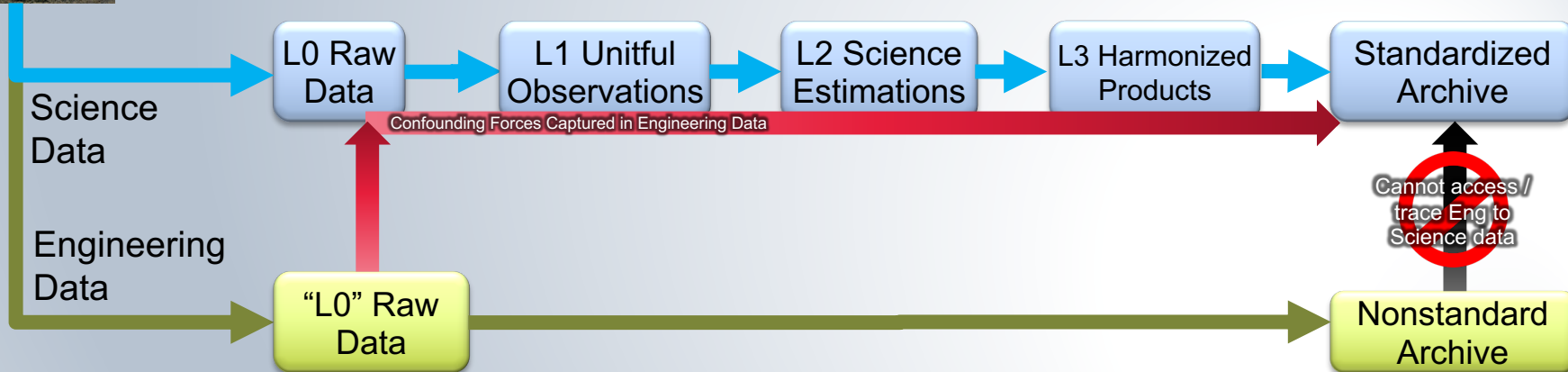
Nonstandard Archive

Motivation: Science Anomaly Situation

- (this really happens!)
- Scientists trace L2/3 anomalies back to L0
- Issue is on instrument! Linked to spacecraft itself
- No worries… scientists will just look at Eng data
- What do you mean it can't be easily harmonized?
- What do you mean I can't easily make graphs?
- Advanced analysis thwarted… spin up Ops effort
- False dichotomy between Science/Eng data maintained

Science Data

Engineering Data

L0 Raw Data → L1 Unitful Observations → L2 Science Estimations → L3 Harmonized Products → Standardized Archive

Confounding Forces Captured in Engineering Data

"L0" Raw Data

Cannot access / trace Eng to Science data

Nonstandard Archive

Motivation: Ops Anomaly Situation

Panic! Streetlight effect forces minimum effort paths

- Look at the eng data we CAN get at
- Check only the most likely relationships
- Confirmational analysis only… exploration hindered
- System interactions are hardest to explore

Engineering Data

Wild & Wooly

"L0" Raw Data
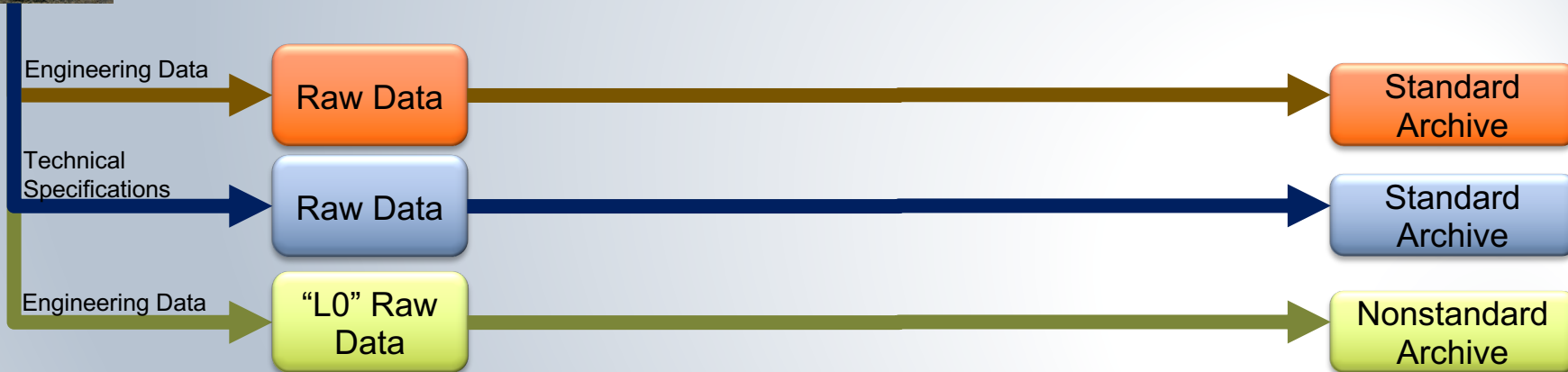
Pathological Formats
Siloed Access
Obscure/missing Docs

Nonstandard Archive

Motivation: Multi-Mission Comparison

Try to compare data from multiple missions or ATLO to Flight
- Totally different archive, formats
- Mature missions no longer have original personnel!
- Requires research project just to harmonize
- Neither archive captured sufficient info for cross-comparison
- May not actually be possible within reasonable time/cost

Engineering Data → Raw Data → Standard Archive

Technical Specifications → Raw Data → Standard Archive

Engineering Data → "L0" Raw Data → Nonstandard Archive

Part of solution already known!

NASA's Analysis Ready Data (ARD) concept: the data we all want & need

Already Preprocessed

Analyst can just start work

Self-Documenting

Easily select and interpret data

Full Context

Event Logs, Known Issues Faulty Channels, Design Intentions, Expert Knowledge

Common Format

Any program can read No special loaders

Data Science Labels
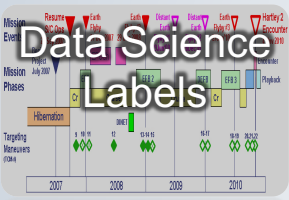
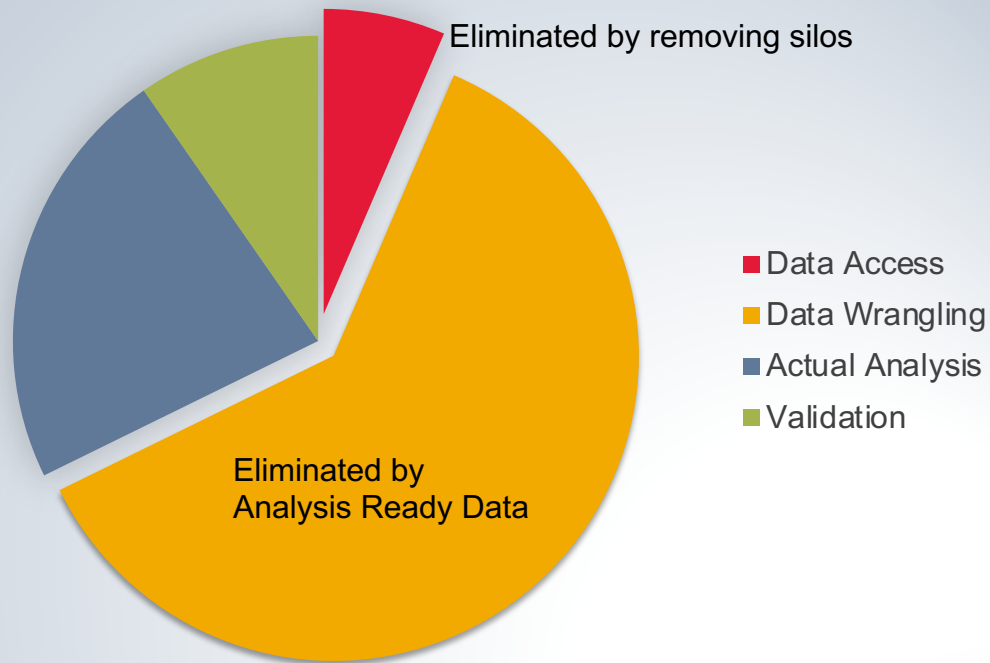Timelines, events, channel information stored as proper labels

Open Database

Simple, searchable, standard, scalable

JPL

Removing Barriers

Data access, format conversion, time-alignment and missing data point management consumes > 60% of time spent on analysis projects.

Analysts must chase SME's to overcome incomplete / esoteric documentation.

Data Science Effort

Eliminated by removing silos

Eliminated by Analysis Ready Data

- Data Access
- Data Wrangling
- Actual Analysis
- Validation

JPL

Continuous Knowledge Capture Tools – Observational Science in Ops!

- Time range annotation (labels)
- What happened when, with categories & notes
- Important patterns & relationships between channels

- Simple, easy, high-level daily questions
- Gives overall daily context as labels
- Ever-growing list of standard categories
- Machine-readable database

- Anomalies should be treated like Hurricanes
- Each has formal name, trajectory, details
- Start & stop time clearly delineated
- Characteristic features specifically highlighted
- Minimize freeform text!

JPL

Delivered solutions – Parallel pipeline for zero disruption
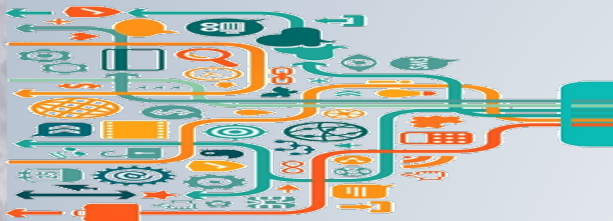
New Pipeline

Established Ops Workflow
(it's made of people!)

Translation Library
- Converts current formats to Data Science formats
- Decodes Ops-specific datatypes (e.g. on-change)

Data Alignment / Harmonization System
- Common time bins, densification (L3 equivalent)
- Merges Metadata

Analytic Database
- On-demand data regridding, filtration
- On-demand data computation results
- On-demand featurization for DS use cases

JPL

Development Lessons Learned

Data Access
➢ Data source APIs optimally need to provide time series for downstream use

Extensibility
➢ Isolate file format (mission) specific code
➢ Fast data point inserts for batch and continuous modes of operation

Data Harmonization
➢ Consider the frame of reference
➢ Beware of features whose data types evolve over time
➢ Revisit repeated value time series

Analysis Ready Data
➢ Store sparse and query to meet the needs of the user
    ➢ Binning rate, interpolation strategies, etc.
➢ Capture accompanying metadata early in spacecraft design
    ➢ Utilize flight rate of occurrence to prioritize creating metadata for active missions
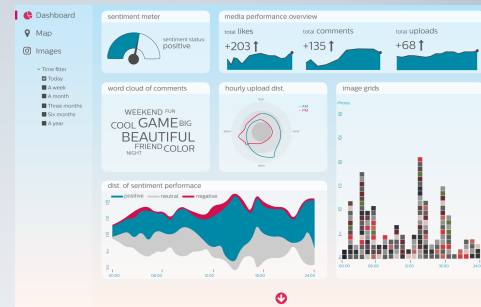
JPL

Tools Enabled by New Pipeline
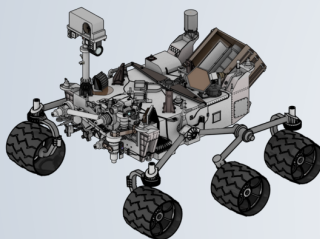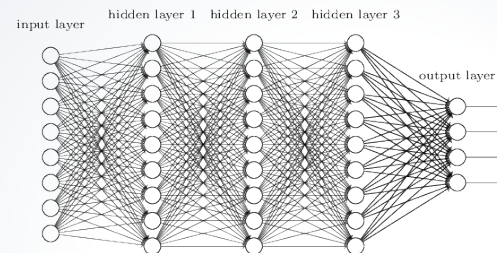
Fast Exploration of ARD

Anomaly detection / Focus of Attention

Interactive Data Dashboarding

Expert Labeling Tools

Multi-Subsystem Trending Tools

Machine Learning Model Infusion